Research status of transportation field based on keyword co-occurrence analysis

Research status of transportation field based on keyword co-occurrence analysis

ZHENG Chun-xiao[1]*, LIN Yao[2], WANG Yang[1], YAN Ya-fei[1], HE XIAO[1], ZHAO Ya-juan[1]

[1]Consulting Service Department National Science Library, CAS Beijing, China

[2]Science and Technology Resources and Technology Exchange Center, China Academy of Transportation Sciences, Beijing, China

*zhengchx@mail.las.ac.cn

Abstract—Transportation is the leading and basic industry in the national economy. It is an indispensable and irreplaceable part of social production and living organization system. The macro analysis of the literature in the field of transportation can provide a new perspective for researchers and managers. This paper uses vosviewer software, based on the keyword co-occurrence relationship of literature, to cluster the conference documents published by the Transportation Research Board of the United States from 2015 to 2019, and analyze in detail eight key areas, including aviation, highway, marine transportation, motor carriers, pedestrians and bicyclists, pipelines, public transportation and railroads. The above analysis reveals the research status of the whole and 8 key sub fields in the field of transportation, so as to provide support for science and technology planning and decision-making in the field of transportation.

1. INTRODUCTION

Transportation, which is the leading and basic industry in the national economy, remains indispensable and irreplaceable in the organization system of social production and living. It appears all-embracing. Even during the same period, the definitions of transportation differ from one another in each country since their transportation situation, development and contemporary focus are not identical; additionally, different scholars even from the same country also have disparate elaborations on transportation engineering due to the differences in their personal perception and understanding [1]. So far, academic circles have still lacked a systematic summary and review of the research status in the whole transportation field.

Literature, a crucial means used by modern researchers for result output and communication, serves as a pool of researcher's study contents, whereas the special literature of conference documents prove superior to journal documents in meeting timeliness. In consequence, the research contents in conference documents are more effective in updating the research status facing researchers in their corresponding field.

This paper systematically summarizes and reviews the research status in the entire transportation field through a keyword co-occurrence analysis of conference documents in this field from the perspective of bibliometrics, in the hope of providing references for scientific and technological innovation, and for planning and decision-making in such field.

2. DATA SOURCES AND SIZE

2.1. Data Sources

Data used for analysis in this paper are drawn from the annual meeting documents released by Transportation Research Board (TRB). All research fields of transportation have been covered by documents submitted during TRB's annual meeting.

2.2. Data Size

2.2.1. Paper classification rules

Conference documents collected are sorted in accordance with the classification rules proposed by TRB[2], covering eight transportation modes, namely aviation, highways, marine transportation, motor carriers, pedestrians and bicyclists, pipelines, public transportation and railroads.

2.2.2. Data trends

A total of 9,013 conference papers were released by TRB from 2015 to 2019. The numbers' trend in different transportation modes is analyzed as shown specifically in Fig.1 and Fig.2

Papers released during 2015 and 2018 enjoyed a stead rise, and peaked in 2018. The following drastic decline in 2019 reduced this year's papers to merely half of the previous year's. Among all the transportation modes highways possess the largest number of documents, only before public transportation.

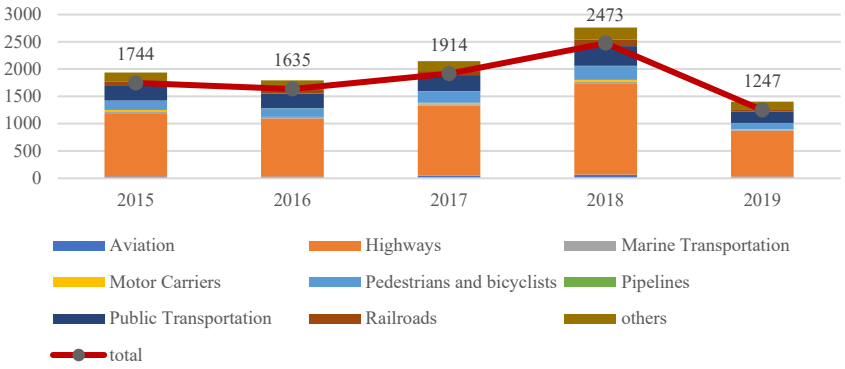

Figure 1: Trends of annual meeting papers released in different transportation modes

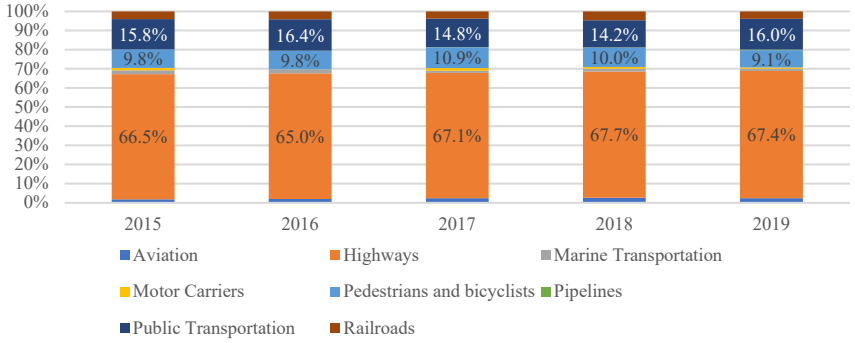

Figure 2: Trends of annual meeting papers released in different transportation modes(percentage)

3. METHODOLOGY

As the data on TRB's documents feature merely five fields, namely author, title, keywords, abstract and publication year, the research status and frontiers in the field of transportation can be fully revealed should keyword co-occurrence analysis, a method used for drawing knowledge maps, be adopted. This paper analyzes keyword co-occurrence in the documents released by TRB by utilizing the knowledge map tool of VOSviewer (visualization of similarities, VOSviewer is a widely used scientific mapping tool, which supports large-scale data processing. It can generate a variety of maps based on Bibliometric relationship, such as keyword co-occurrence diagram, author or journal co citation relationship diagram, etc), which yields the network visualization that can reveal the research status in the field of transportation.

Figure 3: Example of network visualization

In network visualization, an element composed of a node (round spot in the Fig.3) and a text label represents a keyword, whose frequency of co-occurring with other keywords is signified by the size of this node, meaning that a larger node stands for a higher frequency of co-occurrence. The color of a node represents what cluster it belongs to. Different clusters are displayed in different colors. Nodes in the same color represents the documents with the same or similar research orientation. The clustering algorithm applied to VOSviewer is an algorithm for modularity-based community detection in large networks, an optimal version of the community detection algorithm proposed by Newman and Girvan through modularized function weight and parameterized variables[3]. On that basis, Ludo Waltman and Nees Jan van Eck, developers of VOSviewer, have had the clustering function of VOSviewer further optimized by utilizing a smart local moving algorithm (SML)[4], endowing VOSviewer's local clustering with a higher precision.

In network visualization, the thickness of a line linking two nodes is directly proportional to the co-occurrence frequency of the keywords in the same documents, with each keyword represented by a node. A thicker line signifies a higher frequency of co-occurrence. The thickness of a line between two nodes is calculated with the method of full counting.

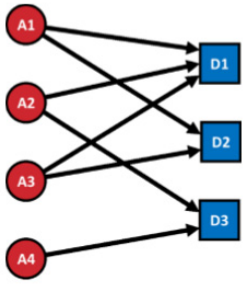

Figure 4: The co-occurrence relationship of four keywords in three documents

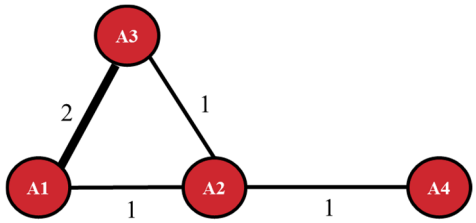

Figure 5: The linking strength between four keywords (line thickness)

The co-occurrence of the keywords A1 and A2 in the sole document D1 means that the linking strength between the two keywords stands at 1; the co-occurrence of A1 and A3 in the two document (D1 and D2) means that their linking strength stands at 2; the co-occurrence of A4 and A2 in the sole document D3 means that their linking strength also stands at 1[5].

Map analysis based on network visualization of the literature keywords in certain field can highlight the current research status in that field.

4. Overall Research Trend

An analysis of the keyword co-occurrence in 9013 documents released by TRB (2015-2019) helps produce a graph as shown in the following figure. The clustering result in the graph demonstrates that 5 major research directions exist in the field of transportation.

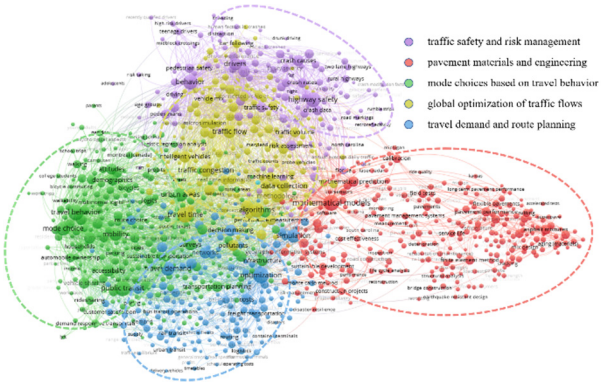

Figure6: Clustering of research in the field of transportation (based on 9013 documents released by TRB)

The purple circle displays research on traffic safety and risk management, in particular highway safety, including cause analysis, detection and forecasting of traffic crashes; the red circle shows research on pavement materials and engineering, including mix design of various mixtures, testing and forecasting of pavement performance, and pavement construction; the green circle demonstrates research on mode choices based on travel behavior, including building choice models after gathering travel behavior by various means, so as to guide traffic control, planning and forecasting; the yellow circle presents research on global optimization of traffic flows, including using various ways to process the traffic flow data accessed from different channels with the aim of planning and forecasting traffic flows; the blue circle concerns research on traffic mode choices based on different travel demand, including study on the impacts that route planning, traffic costs and service quality impose on mode choices.

5. Research Trends in Key Fields

A deep analysis is conducted centering on the 8 types of transportation modes classified by TRB. The clustering and intensity of the keyword co-occurrence serve as the basis for analyzing and exploring the research status in each transportation mode.

5.1. Aviation

The following figure shows the cluster analysis of TRB's documents regarding the transportation mode of aviation.

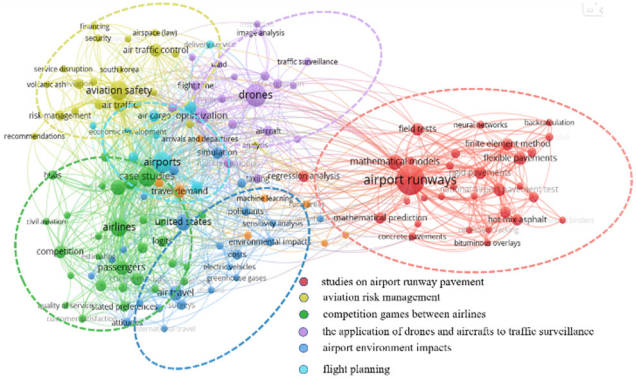

Figure 7: Clustering of research on aviation

In the red circle is analysis of flexible pavement and concrete pavement of airport runway by adopting finite element method, back calculation and neural network; in the yellow circle is research on aviation safety, risk management and air traffic control; in the purple circle is research on traffic surveillance with the aid of drones and aircrafts; in the green circle is research on airlines' competition patterns in each country from the perspective of airport operation, service quality, and customer satisfaction; in the dark blue circle is research concentrating on the environmental impacts imposed by the emissions of greenhouse gases and pollutants during airport transportation, including the use of electric vehicles for airport transportation and the fuel options for airport fleets; in the light blue circle is research focusing on using network analysis to plan and forecast air cargo and flight delays.

5.2. Highways

The following figure shows the cluster analysis of TRB's documents concerning the transportation mode of highways. Four major research directions exist in highways, namely:

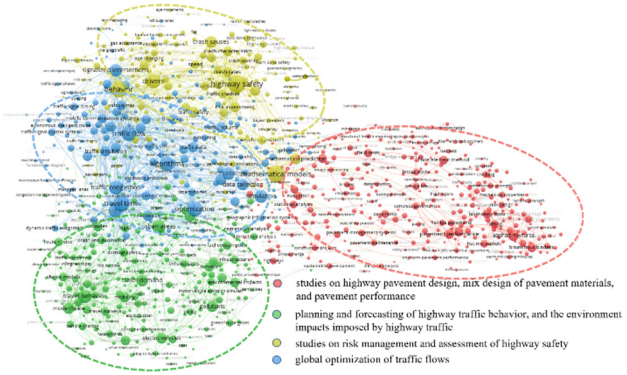

Figure 8: Clustering of research on highways

Red circle: research on the design of highway pavement(including bridge decks), pavement base course and highway pavement material mix, and on the performance of different pavements, including reclaimed asphalt, asphalt and flexible pavement. Research on highway pavement design focuses on

adopting mechanistic empirical pavement design to build prototypes, so as to forecast the deterioration of pavement performance; research on mix design comprises that of asphalt mixtures, hot mix asphalt, bituminous binder and warm mix paving mixtures; research on pavement performance centers on pavement mechanical properties, including experimenting, testing and modeling in terms of compressive strength, earthquake resistance, durability, cracking, the occurrence and control of rutting, and service life.

Yellow circle: research on the risk assessment and management of highway safety through mathematical models. Models are built by using the crash data (for instance, crash rates, fatalities and crash severity) on traffic crashes, including pedestrian-vehicle and bicycle crashes. Crash characteristics extracted from the models and crash causes revealed are used to guide highway design (for instance, roundabout design), the protection of pedestrian safety, and the management of drivers, in particular teenage drivers, in hope of reducing crash incidence and safeguarding traffic safety.

Blue circle: research conducted centering on the global optimization of traffic flows. To be specific, it includes: research on the coordination of traffic signals, namely building optimal models of coordinated signal timing parameters in the light of traffic signal control systems and traffic signal timing, so as to enhance the coordination of traffic signal timing parameters; research on designing automated highway systems or intelligent highway systems, namely the application of autonomous intelligent cruise control, adaptive control, mobile communication systems, connected vehicles, and intelligence vehicles to the design of automated highway systems or intelligent highway systems; research on highway traffic control, namely locating the optimum strategy for highway traffic control by applying traffic simulation modeling to the analysis of various traffic flows (including autonomous vehicle and pedestrian-vehicle mixed flows), so as to minimizing the incidence of traffic congestion; access and analysis of traffic flow data, namely utilizing modern information technologies, such as multivariate data fusion and global positioning systems, to analyze the traffic flow data accessed by means of multiple detection methods, including detection and identification systems, probe vehicles and sensors, before putting analysis results into the use of research on highway capacity and the design of intelligent highway systems.

Green circle: research on the planning and forecasting of highway travel behavior, and on the environmental impacts imposed by highway traffic. In terms of the former aspect, surveys of public travel demand and travel behavior are conducted in order to analyze people's mode and route choices in highways. Traffic choice models built are used for studying the causes underlying and the influences resulting from different choice models, with the aim of planning and forecasting people's highway travel behavior. Among the highway traffic mode are sharing vehicles, electric vehicles and public transportation, while factors affecting route choices contain trip length, vehicle miles of travel, and family activities. In terms of the latter aspect, studies concentrate on what influences the application of electric vehicles and plug-in hybrid vehicles to highway traffic will have on the reduction of carbon dioxide emissions, which specifically includes analyzing the benefits that a quantitative vehicle electrification strategy creates to energy consumption, the impacts that power sources of electric vehicles exert on greenhouse gas emissions, and the factors that affect the energy efficiency of vehicles.

5.3. Marine Transportation

The following figure displays the cluster analysis of TRB's documents referring to the transportation mode of marine transportation.

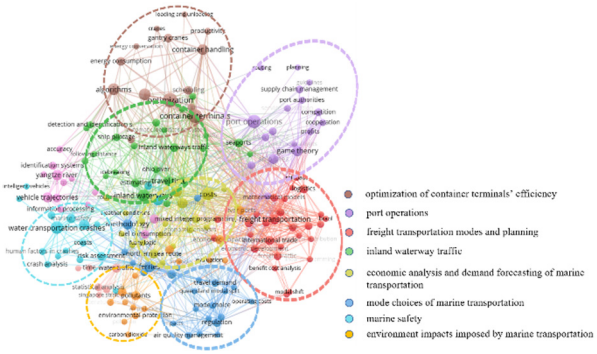

Figure 9: Clustering of research on marine transportation

Brown circle: optimization of container terminals' efficiency. Various algorithms are adopted to improve the productivity of container terminals and container ships, and to minimize energy consumption.

Red circle: modes and planning of freight transportation in international trade. Among the freight transportation modes are container traffic, railroad intermodal transportation and multimodal facilities. In contrast, freight transportation planning contains the benefit-cost analysis of logistics by using the mathematical model of linear programming.

Green circle: research related to inland waterway traffic, including ship pilotage in inland waterways, ice-breaking, dredging and locks of waterways.

Deep blue circle: research on mode choices of marine transportation, including study on the effects that prices have on mode choices based on modal split models, and the effects that routes, travel demand, air quality management, special sulphur emission regulation, and operation costs impose on mode choices.

Yellow circle: economic analysis and demand forecasting of marine transportation, which mainly comprises the demand forecasting and scheduling optimization (including routes and schedules) of liner shipping.

Purple circle: research on port operations, including study on the application of game theory to port cooperation and competition, and to supply chain management, as well as study on the impacts that port capacity and profits exert on port operations.

Light blue circle: research on marine safety, including simulating water transportation crashes to facilitate risk analysis, identifying crash causes and human factors in crashes underlying water transportation crashes, and determining the incidence of water transportation crashes through risk assessment of marine transportation.

Orange circle: environmental impacts imposed by pollutants from marine transportation, for instance, air pollution and carbon dioxide.

5.4. Motor Carriers

The following figure displays the cluster analysis of TRB's documents involving the transportation mode of motor carriers.

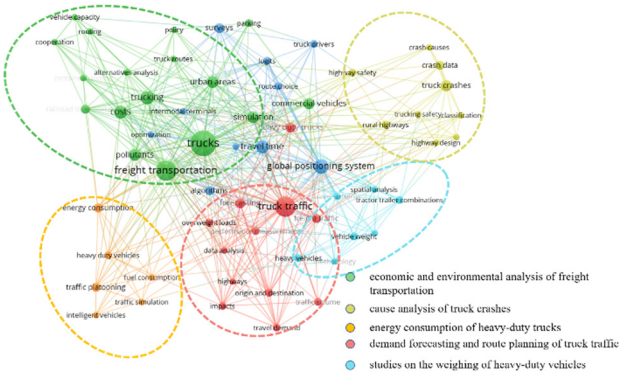

Figure 10: Clustering of research on motor carriers

Green circle: economic and environmental analysis of trucks used for freight transportation, including the optimization of truck routes, scheduling of container handling and vehicle capacity based on simulation models, and also including the reduction of transportation costs and pollutant emissions through alternatives analysis.

Yellow circle: crash analysis of highway safety, concentrating on utilizing regression analysis and crash data to explore the causes and classifications of truck crashes, and the effects resulting from highway design.

Orange circle: research on the energy consumption of heavy-duty vehicles, mainly regarding reducing energy consumption and transportation costs by installing or improving vehicle-mounted systems, and by developing and employing the technologies of traffic platooning and traffic simulation.

Red circle: research on truck traffic, including route planning between origins and destinations for freight traffic, demand forecasting of transportation through data analysis, and the analysis of impacts that heavy duty trucks, in particular those under overweight loads, impose on highways.

Light blue circle: research on the weighing of heavy vehicles, including using the spatial analysis methodology and weigh-in-motion systems to analyze the travel patterns and weight distributions of tractor trailer combinations on the basis of GPS database, so as to improve the operation of transportation network.

5.5. Pedestrians and Bicyclists

The following figure shows the cluster analysis of TRB's documents concerning the transportation mode of pedestrians and bicyclists.

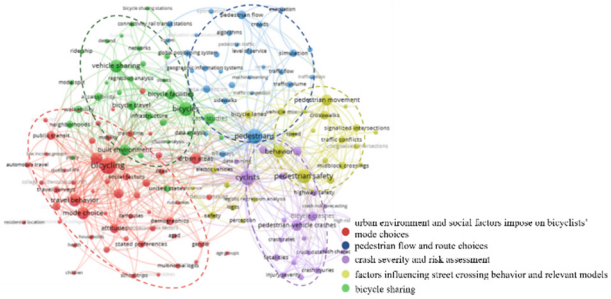

Figure 11: Clustering of research on pedestrians and bicyclists

The red circle displays studies focusing on the impacts that urban environment and social factors impose on the mode choices of bicycle travel. Among the factors affecting urban environment are air pollution, city planning, built environment, bus accessibility, and infrastructure for non-motorized transport; social factors contain gender differences, trip length, comfort level and travel time. Meanwhile, sociodemographic and contingent valuation methods, the latter belonging to stated

preference methods, are employed to classify bicyclists before analyzing the related factors that affect choices of travel routes.

The blue circle shows studies concentrating on pedestrian flow models, sidewalks and route choices. The models built include two-way pedestrian, pedestrian evacuation, pedestrian level of service, walkway choice, and social force models for pedestrian flows with real psychological behavior. Studies are also performed regarding the characteristics of non-uniform flows, pedestrian counter-flows with left- or right-moving preference, distribution of street space, and pedestrians' psychological behavior of collision avoidance.

The purple circle demonstrates assessment of the factors affecting injury severity, and of traffic crash risks. The factors specifically contain weather conditions, motor vehicle crashes, street crossing safety, urban environment, and road environment. The data mining of traffic crashes involving pedestrians serves as a basis for modeling, including models of pedestrian collision at signalized intersections, forecasting of traffic crashes involving pedestrians at signalized intersections, collision frequency and ratio based on Bayesian approach, multinomial logit analysis of traffic crashes, and space-time integrated models of injury severity in pedestrian-vehicle crashes.

The yellow circle displays research on the factors influencing street crossing behavior in crosswalks and at signalized intersections, and on related models. Among factors affecting street crossing behavior are gender of crossing groups, perception of vehicles and other pedestrians, pedestrian movement, traffic signs along pedestrian crossings, and crash risk of right-turn vehicle movement. Additionally, models built for deep analysis include social force models for pedestrian behavior, pedestrian violations and their change in trajectories , pedestrian dynamics, theory on pedestrian gap acceptance, pedestrian decisions on street crossing, and intensity models of pedestrian crossing risk.

The green circle concerns research centering on factors affecting bicycle travel, among which the major ones comprise weather conditions, built environment, travel time, urban morphology, locations of public bicycle stations, gender, age, and suspension and accessibility of public transportation. Furthermore, models of forecasting the traffic flow transfer probability regarding shared bicycles, and choice models of destination preference are built on the basis of the abovementioned factors.

5.6. Pipelines

A total of 13 documents released by TRB concentrate on the transportation mode of pipelines. Among the top 5 research topics in pipelines are maintenance and preservation (5 documents), bridges and other structures (2 documents), safety and human factors (2 documents), planning and forecasting (2 documents), and materials (2 documents).

5.7. Public Transportation

Fig.12 shows the cluster analysis of TRB's documents centering on the transportation mode of public transportation.

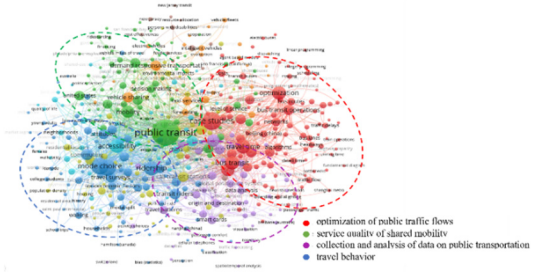

Figure 12: Clustering of research on public transportation

The green circle focuses on transport sharing — including ride sharing and ridesourcing — and quality of service, including performance evaluation and assessment, and service disruption.

The red circle refers to the application of mathematical models, produced though mathematical algorithms, to the simulation analysis of traffic flow, travel time, traffic capacity and traffic network, so as to resolve realistic problems ranging from traffic congestion to traffic delays.

The blue circle centers on travel behavior, containing models of traffic mode choices, traffic modal split, specific travel behavior, and travel surveys conducted to analyze customer satisfaction and passenger attitudes.

The purple circle displays research on applying the methodology of cluster, spatial and statistical analysis to data analysis, data collection, automatic data collection systems.

5.8. Railroads

The cluster analysis of TRB's documents concerning the transportation mode of railroads is as shown in Figure 13.

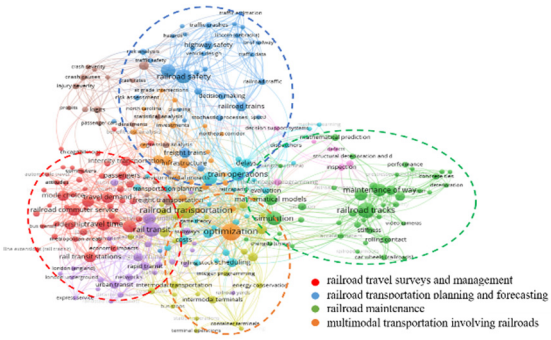

Figure 13: Clustering of research on railroads

The blue circle displays the analysis of traffic data on railroads. Estimations of the traffic flow trend contribute to having the running scheduling of railroads arranged and planned properly. Additionally, the risk of traffic safety is assessed in studies on freight trains, tank cars and railroad trains in BNSF Railway, America (Nebraska and Texas), Canada and Taiwan Province, China.

The red circle concentrates on high-speed rail systems, including high-speed rail and subways. Travel surveys are conducted to study commuters' and passengers' satisfaction in order to analyze the effects that the quality of service in railroad transportation has on passengers' travel behavior. Moreover, analysis is made regarding the economic impacts that travel demand, travel time and traffic accessibility impose on economic development of railroad transportation.

The orange circle concerns in-depth research chiefly on China's and American railroad transportation, two countries which are respectively ranked first and second worldwide in their total mileage of railways. Three particular aspects are included. One is the field of freight transportation, for instance, rail intermodal service and container handling; another refers to travel surveys of various indicators suggesting passenger experience, including passenger comfort and passenger volume; still another involves analyzing and forecasting the incidence of railroad crashes by adopting the statistical method of regression analysis, for instance, injury severity, crash causes, crash severity, crash rates, crash risk forecasting and derailments.

The green circle centers on railway maintenance. Optimal models of railway maintenance planning are built on the basis of integer programming, and meanwhile ultrasonic detectors are utilized to carry out non-destructive inspections of the potential defects in steel rails, so that causes underlying railroad diseases can be identified. The causes range from repetitive vibration loads to stiffness, impact loads, durability and structural deterioration.

6. CONCLUSION

In light of the data on the documents released by TRB (2015-2019), this paper detailedly analyzes the research status in eight key fields, namely aviation, highways, marine transportation, motor carriers,

pedestrians and bicyclists, pipelines, public transportation and railroads. The analysis reveals that research in the field of transportation centers on the following eight aspects:

6.1. Studies on pavements
Studies in this aspect mainly contain pavement design, mix design of pavement materials, pavement performance testing and improvement, and the application of flexible pavement to airport runways.

6.2. Planning and forecasting of transportation
Traffic data accessed through various monitoring methods are used for studies on optimizing traffic flows so as to realize transportation planning and demand forecasting. Studies in this aspect mainly include the transportation planning and demand forecasting of flights, highway traffic, freight traffic, marine transportation and railway transportation.

6.3. Risk management of transportation
Studies in this regard chiefly comprise aviation risk management, crash analysis, risk management and assessment of highway safety, crash severity and risk assessment of pedestrians and bicyclists, risk analysis of railroad crashes, and marine safety.

6.4. Traffic terminal operations
Studies in this aspect center on competition games between airlines, port and terminal operations, and optimization of container terminals' efficiency.

6.5. Mode choices of transportation
In this aspect various survey methods (for instance, stated preference) are adopted to research different travel behavior, which serves as a basis for modeling transportation mode choices. Among the factors affecting modeling are O-D (origin and destination), service quality of transportation, and travel time.
Models of transportation mode choices during delivery are also under the influence of energy consumption of transportation vehicles, and weighing of heavy vehicles.

6.6. Environment impacts imposed by transportation
Studies in this regard mainly contain analyzing environment impacts imposed by highways, marine transportation and airport transportation, exploring freight transportation environment, and studying environmental improvement thanks to bicyclists.

6.7. Economic impacts imposed by transportation
Studies in this aspect focus on economic analysis of marine and freight transportation, and on the impacts that railroads have on economic development.

6.8. Disease detection, repairs and maintenance of pavement, steel rails and pipelines
Worldwide transportation is ushering in a new developmental stage in history, in which new forms of business, new technologies and new models continually replace the previous ones. Data used for analysis in this paper are all from the documents released by TRB. Undoubtedly, the necessary period of time from undertaking this research to writing and publishing this paper engenders a time lag, which might affect the judgment on research status. Despite that, should the keyword co-occurrence analysis, a method used for drawing knowledge maps, be adopted, the research status in the field of transportation can still be objectively reflected due to the shorter publishing period of the conference documents released by TRB compared with journal documents.

REFERENCES

[1] Editorial Department of China Journal of Highway and Transport,"Review on China's Traffic Engineering Reaearch Progress :2016," China Journal of Hignway and Transport,vol. 29(06),pp.1-161,2016.

[2] TRB. 2020. TRB Publications by Subject.: http://www.trb.org/Publications/PubsTRBPublicationsbySubject.aspx. [2020-03-15].
[3] Newman M E J ,Girvan M.“Finding and Evaluating Community Structure in Networks”. Physical Review E, vol.69(2 Pt 2):026113. 2004.
[4] Waltman L, Eck N J V .“A smart local moving algorithm for large-scale modularity-based community detection,” The European Physical Journal B, vol.86(11),pp.1-14,2013.
[5] VOSviewer.2017.VOSviewer Manual. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.6.pdf.[2020-03-15]